

Figure 1

Minimum signal peptide score	false positive rate	false negative rate	proba(0.1)	proba(0.2)
3.5	0.121	0.036	0.467	0.664
4	0.096	0.06	0.519	0.708
4.5	0.078	0.079	0.565	0.745
5	0.062	0.098	0.615	0.782
5.5	0.05	0.127	0.659	0.813
6	0.04	0.163	0.694	0.836
6.5	0.033	0.202	0.725	0.855
7	0.025	0.248	0.763	0.878
7.5	0.021	0.304	0.78	0.889
8	0.015	0.368	0.816	0.909
8.5	0.012	0.418	0.836	0.92
9	0.009	0.512	0.856	0.93
9.5	0.007	0.581	0.863	0.934
10	0.006	0.679	0.835	0.919

Figure 2

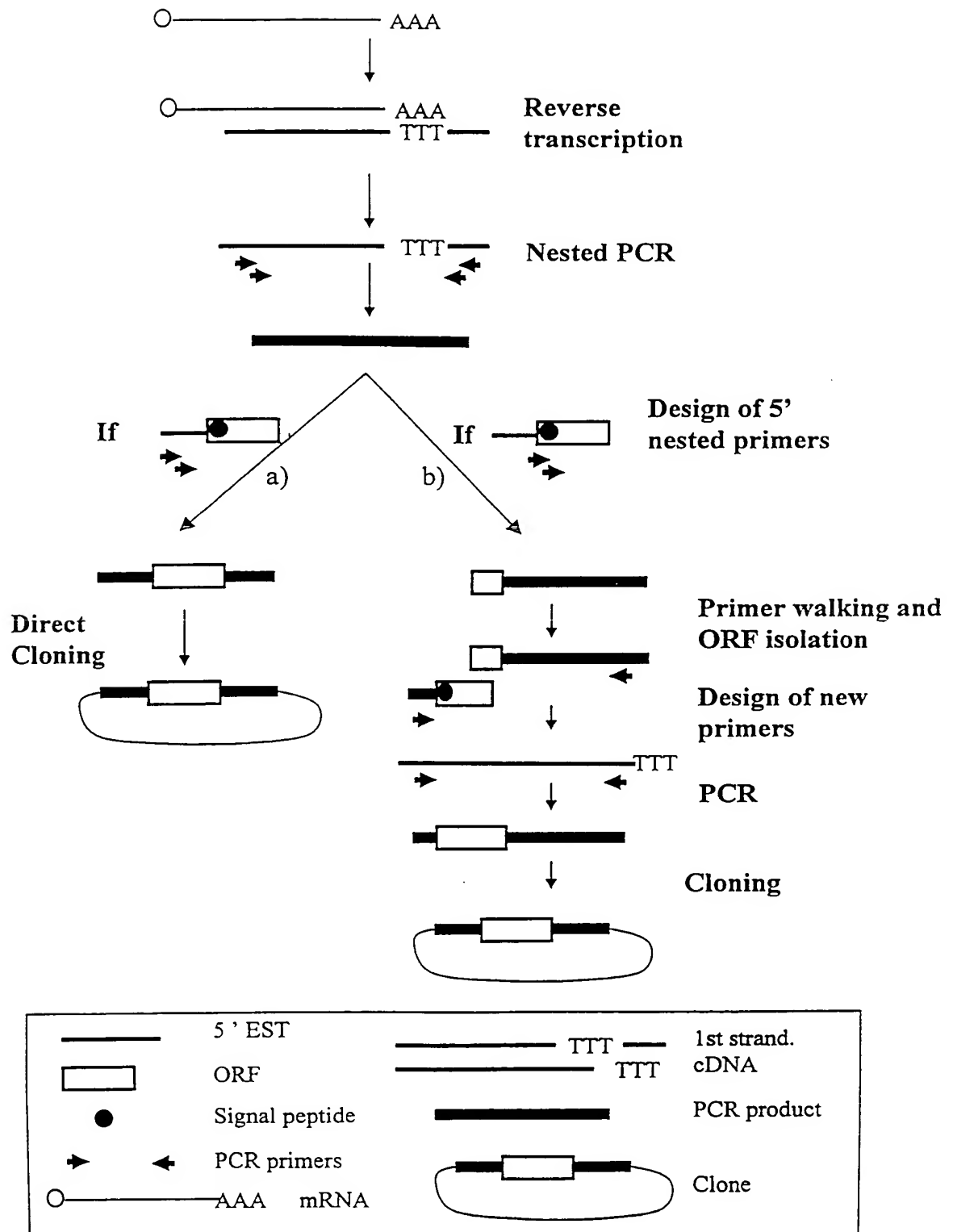


Figure 3

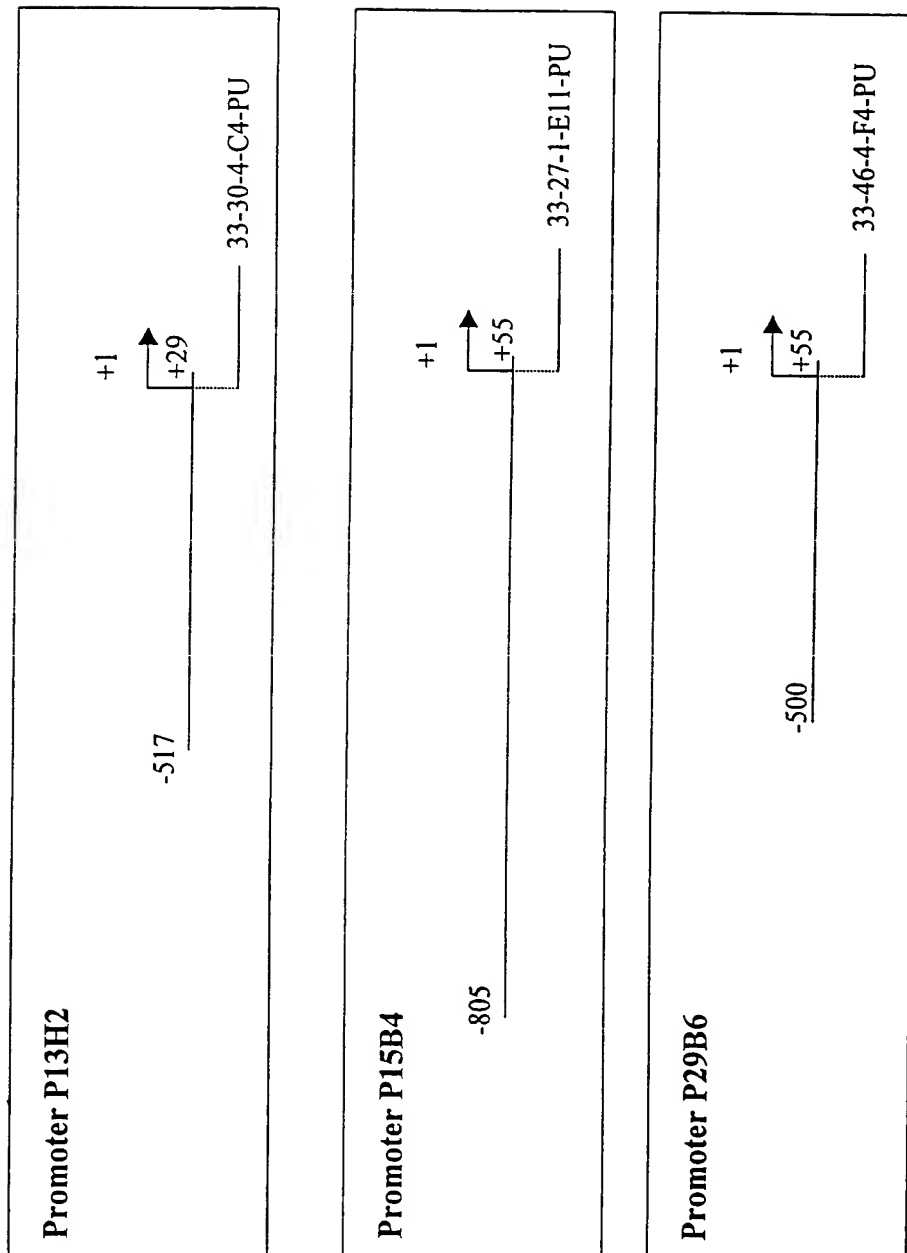


Figure 4

Description of Transcription Factor Binding Sites Present on Promoters Isolated From SignalTag Sequences					
Promoter sequence P13H2 (546 bp):					
Matrix	Position	Orientation	Score	Length	Sequence
CMYB 01	-502	+	0.983	9	TGTCAGTTG
MYOD Q6	-501	-	0.961	10	CCCAACTGAC
S8 01	-444	-	0.960	11	AATAGAATTAG
S8 01	-425	+	0.966	11	AACTAAATTAG
DELTAEF1 01	-390	-	0.960	11	GCACACCTCAG
GATA C	-364	-	0.964	11	AGATAAATCCA
CMYB 01	-349	+	0.958	9	CTTCAGTTG
GATA1 02	-343	+	0.959	14	TTGTAGATAGGACA
GATA C	-339	+	0.953	11	AGATAGGACAT
TAL1ALPHAE47 01	-235	+	0.973	16	CATAACAGATGGTAAG
TAL1BETAE47 01	-235	+	0.983	16	CATAACAGATGGTAAG
TAL1BETAIF2 01	-235	+	0.978	16	CATAACAGATGGTAAG
MYOD Q6	-232	-	0.954	10	ACCATCTGTT
GATA1 04	-217	-	0.953	13	TCAAGATAAAGTA
IK1 01	-126	+	0.963	13	AGTTGGGAATTCC
IK2 01	-126	+	0.985	12	AGTTGGGAATTCC
CREL 01	-123	+	0.962	10	TGGGAATTCC
GATA1 02	-96	+	0.950	14	TCAGTGATATGGCA
SRY 02	-41	-	0.951	12	TAAAACAAAACA
E2F 02	-33	+	0.957	8	TTTAGCGC
MZF1 01	-5	-	0.975	8	TGAGGGGA
Promoter sequence P15B4 (861bp) :					
Matrix	Position	Orientation	Score	Length	Sequence
NFY Q6	-748	-	0.956	11	GGACCAATCAT
MZF1 01	-738	+	0.962	8	CCTGGGGA
CMYB 01	-684	+	0.994	9	TGACCGTTG
VMYB 02	-682	-	0.985	9	TCCAACGGT
STAT 01	-673	+	0.968	9	TTCTTGGA
STAT 01	-673	-	0.951	9	TTCCAGGA
MZF1 01	-556	-	0.956	8	TTGGGGGA
IK2 01	-451	+	0.965	12	GAATGGGATTTCC
MZF1 01	-424	+	0.986	8	AGAGGGGA
SRY 02	-398	-	0.955	12	GAAAACAAAACA
MZF1 01	-216	+	0.960	8	GAAGGGGA
MYOD Q6	-190	+	0.981	10	AGCATCTGCC
DELTAEF1 01	-176	+	0.958	11	TCCCACCTTCC
S8 01	5	-	0.992	11	GAGGCAATTAT
MZF1 01	16	-	0.986	8	AGAGGGGA
Promoter sequence P29B6 (555 bp) :					
Matrix	Position	Orientation	Score	Length	Sequence
ARNT 01	-311	+	0.964	16	GGACTCACGTGCTGCT
NMYC 01	-309	+	0.965	12	ACTCACGTGCTG
USF 01	-309	+	0.985	12	ACTCACGTGCTG
USF 01	-309	-	0.985	12	CAGCACGTGAGT
NMYC 01	-309	-	0.956	12	CAGCACGTGAGT
MYCMAX 02	-309	-	0.972	12	CAGCACGTGAGT
USF C	-307	+	0.997	8	TCACGTGC
USF C	-307	-	0.991	8	GCACGTGA
MZF1 01	-292	-	0.968	8	CATGGGGA
ELK1 02	-105	+	0.963	14	CTCTCCGGAAGCCT
CETS1P54 01	-102	+	0.974	10	TCCGGAAGCC
AP1 Q4	-42	-	0.963	11	AGTGACTGAAC
AP1FJ Q2	-42	-	0.961	11	AGTGACTGAAC
PADS C	45	+	1.000	9	TGTGGTCTC

Figure 5

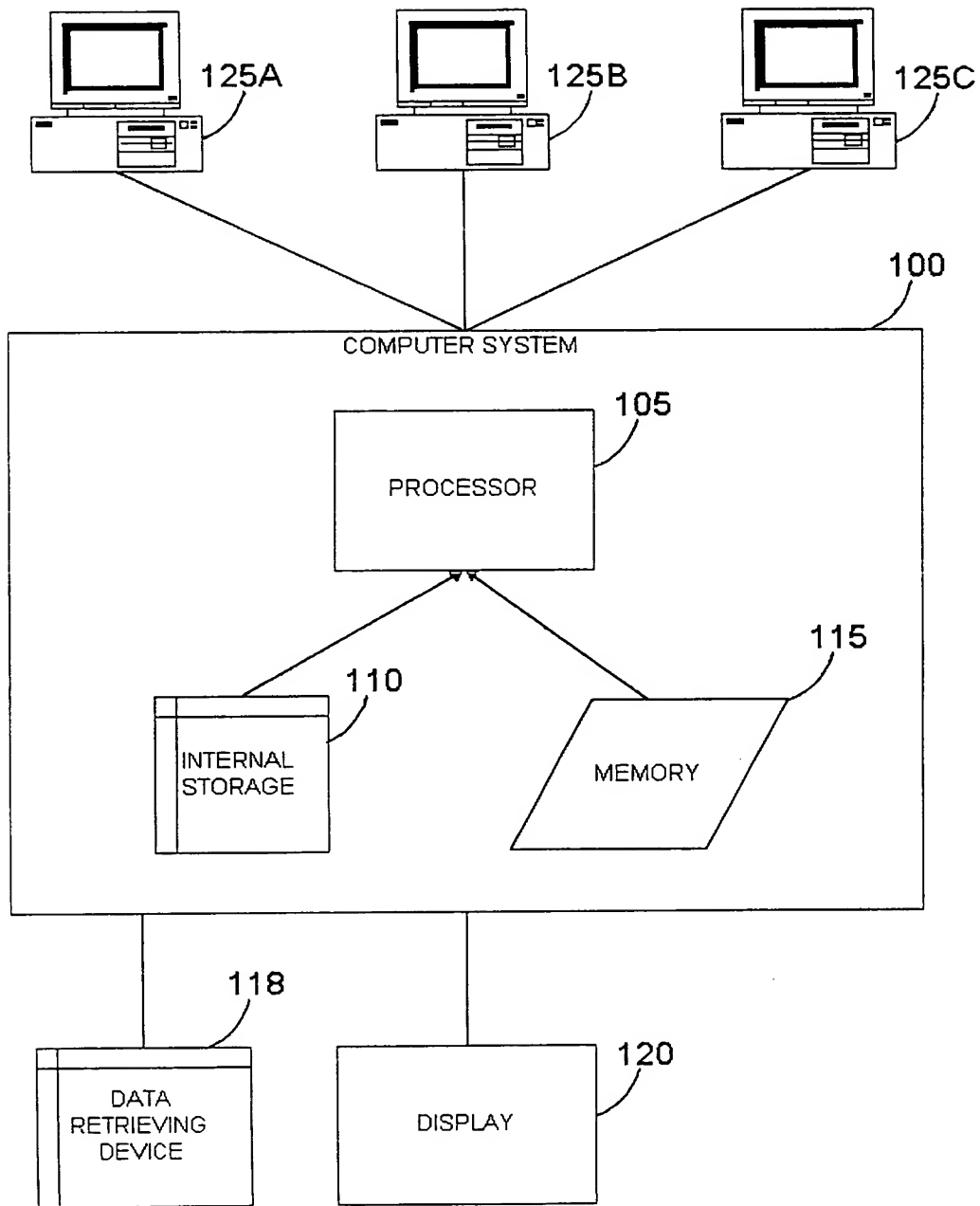


FIGURE 6

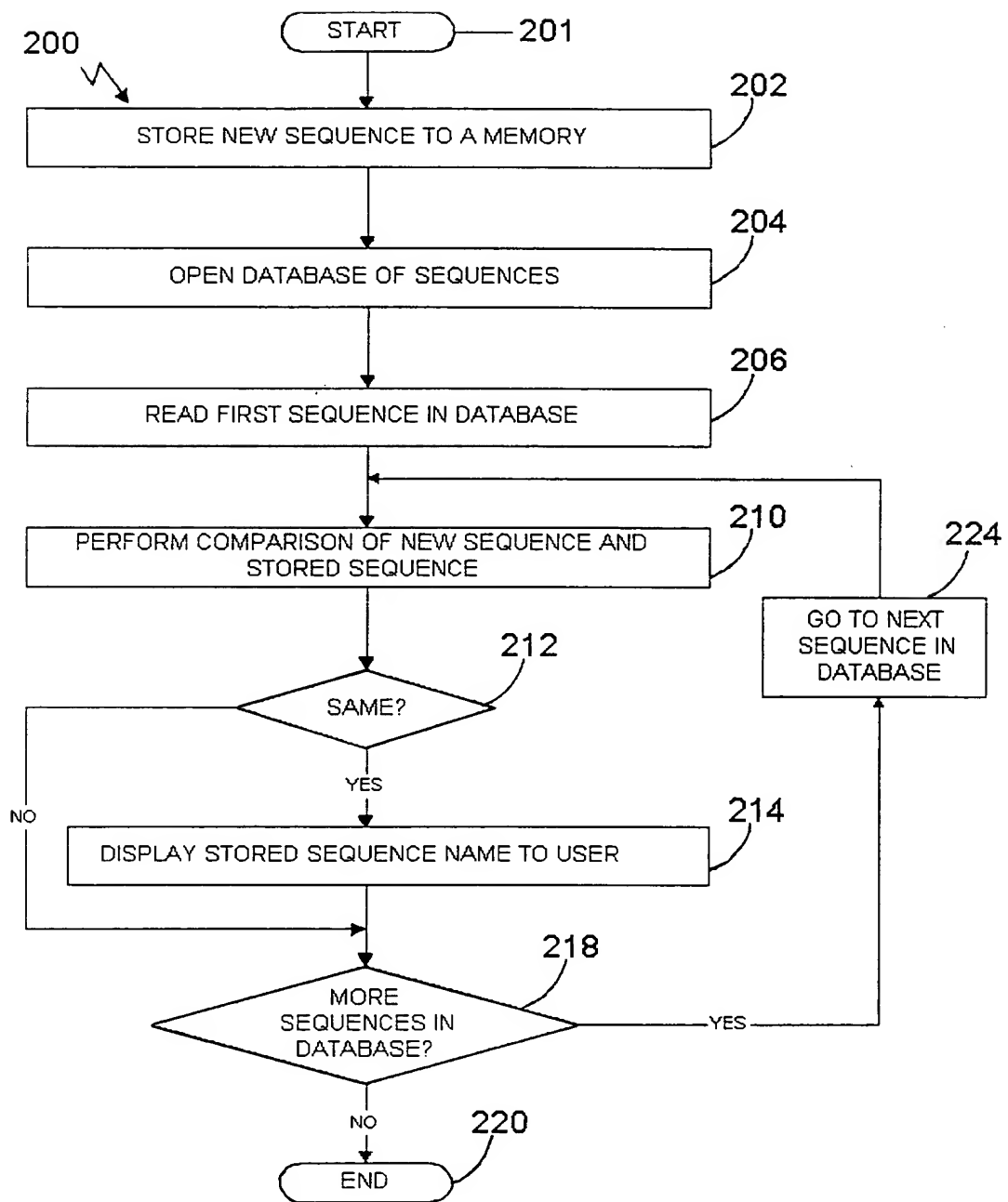


FIGURE 7

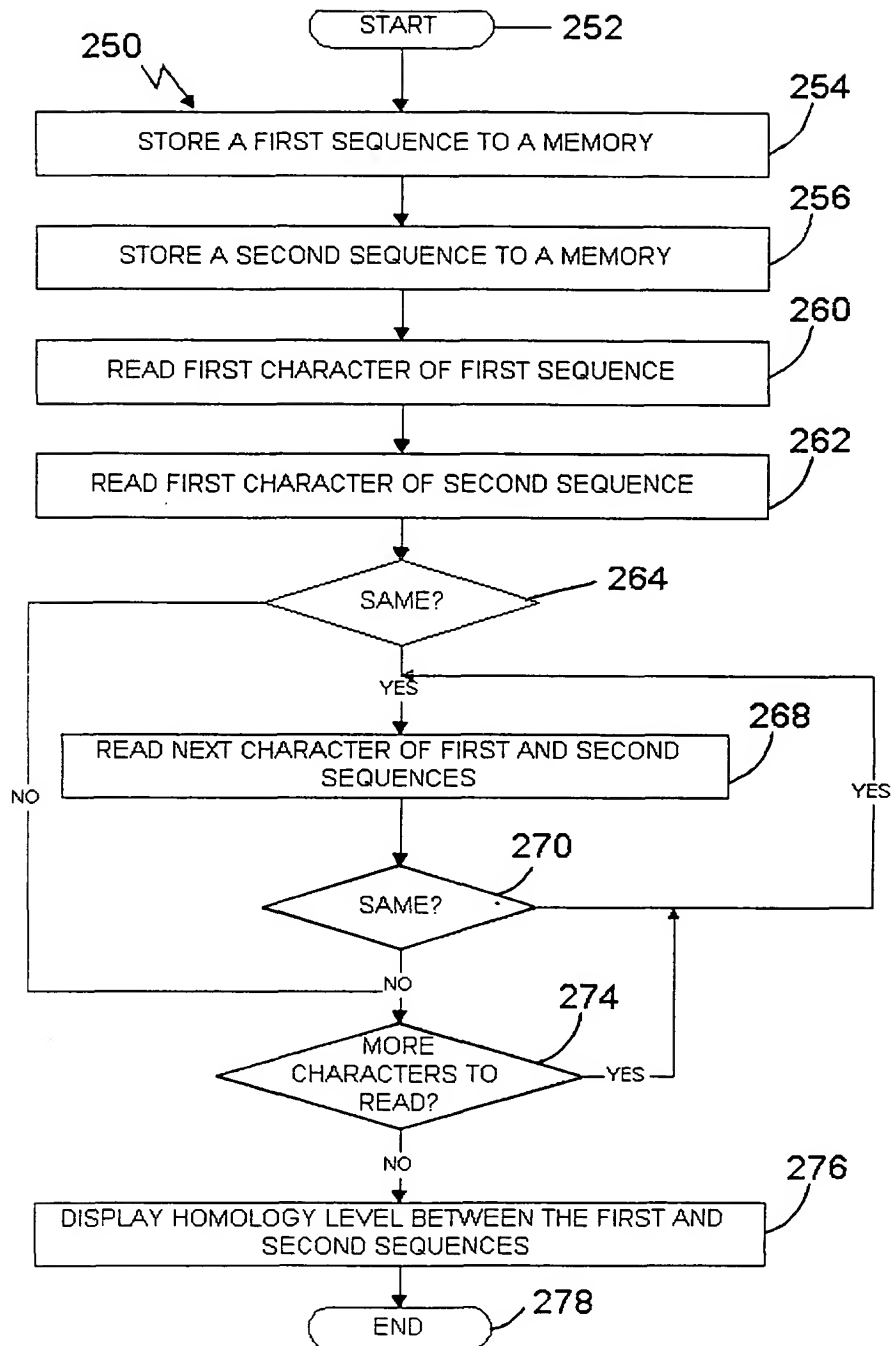


FIGURE 8

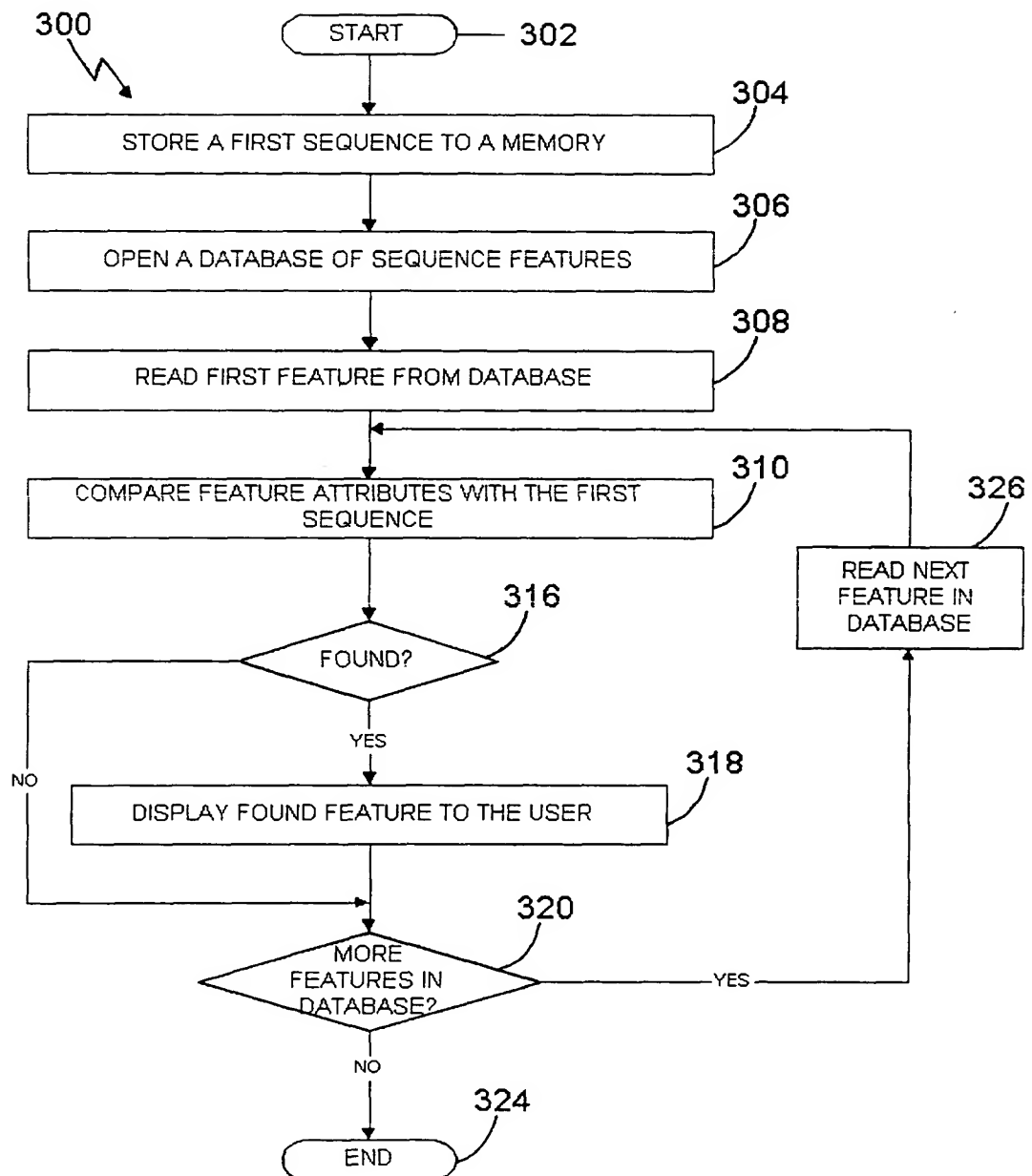


FIGURE 9

Search characteristic		Selection Characteristics		
Step	Program	Strand	Parameters	Identity (%) Length (bp) Comments
miscellaneous	FASTA	both	-	90 15
tRNA	FASTA	both	-	80 60
rRNA	BLASTN	both	S=108	80 40
mtRNA	BLASTN	both	S=108	80 40
Prokaryotic	BLASTN	both	S=144	90 40
Fungal	BLASTN	both	S=144	90 40
Alu	BLASTN	both	S=72, B=5	70 40 max 5 matches, masking
L1	BLASTN	both	S=72, B=5	70 40 max 5 matches, masking
Repeats	BLASTN	both	S=72	70 40 masking
PolyA	BLAST2N	top	W=6, S=10, E=1000, N=-12	90 10 in the last 100 nucleotides
Polyadenylation signal	-	top	AATAAA allowing 1 mismatch	in the 50 nucleotides preceding the 5' end of the polyA
Vertebrate	BLASTN then FASTA	both	-	90 then 70 30 first BLASTN and then FASTA on matching sequences
ESTs	BLAST2N	both	-	90 30
Geneseq	BLASTN	both	W=8, B=10	90 30
ORF	BLASTP	top	W=8, B=10	- on ORF proteins, max 10 matches
Proteins	BLASTX	top	E = 0.001	70 30

Figure 10